

Systems Biology, Integrative Biology, Predictive Biology Commentary

Edison T. Liu*

Genome Institute of Singapore
60 Biopolis Street
02-01 Genome, 138672
Singapore

The pursuit of science is based on empiricism enabled by advancing technologies. Theories then emerge that unify our understanding of natural order. The trinity of experimentation, theory, and technology that underlies the scientific method cannot be placed into a hierarchy of importance. Theoretical fields quickly lose substance in the absence of robust experimentation, and theoretical conundrums cannot be resolved without experimentation using advanced technologies.

Physics is perhaps the most mature of the experimental sciences in exploiting the interface between theory, experimentation, and technology. Quantum and relativity theories revolutionized physics and engineering, but high-energy physics is a discipline made possible by the technology of the linear accelerator. The advantage of physics over other experimental sciences is the ability to use first principles to calculate more complex models of the physical world. These models are then used to predict unforeseen physical events. Fundamental to this is the ability to precisely quantify physical observations and to render them mathematically transformable.

This progression from observation, to theory, to prediction has been the driving force in humanity's fascination with science. Not only do we have an inherent appetite for knowledge that motivates us to explore, but we also have a desire to extract reason from perceived chaos in order to construct models that can predict our future.

In this past century, biology has transformed from a primarily empiric and observational field to one based on fundamental principles. The foundations for this change may be attributed to a great degree to the discovery of the true structure of DNA, and followed by the recent ability to sequence entire genomes. We are now poised to advance further into predictive theory with the new discipline of systems biology.

Systems biology can be different things to different people, but there is a common understanding that this discipline seeks to explain biologic phenomenon, not on a gene-by-gene basis, but through the net interactions of all cellular and biochemical components within a cell or organism. Operationally, systems biology requires the ability to digitalize biological output so that it can be computed, the computational power to analyze comprehensive and massive datasets, and the capacity to integrate heterogeneous data into a usable knowledge format. Thus, systems biology can be described as “integrative biology” with the ultimate goal of being able to predict *de novo* biological outcomes given the

list of the components involved. An example of this approach is seen in the work by Davidson and colleagues. After identifying the important components for developmental specification for sea urchin endomesodermal differentiation (Davidson et al., 2002a), they defined the genetic network for endomesodermal specification on a genome-wide scale (Davidson et al., 2002b; Bolouri and Davidson, 2003). By describing the network as a precise functional model mapped by a control diagram, they advanced the prospects of forward engineering of a complex biological process. In this aspect, systems biology is also “predictive biology.”

The Genome Institute of Singapore, founded in 2001, is focused on integrative and systems approaches to resolving biological questions of medical importance. We seek to integrate biological and analytical systems to generate large-scale and precise data about dynamic cellular processes (Wei et al., 2005; Peng et al., 2005; Lin et al., 2004). We, and others, believe that this approach not only can provide an unbiased and complete dataset about a biological system, but when analyzed appropriately, can give insights as to the interconnections of molecular components and the hierarchy of importance of the often large number of involved molecules and pathways (Liu, 2004; Hood et al., 2004). To this end, we have focused on the dynamics of the transcriptome using genomic tools (Ng et al., 2005; Ruan et al., 2004). This strategy was selected in large part because the technologies for precisely assessing the transcriptome in a highly multiplexed fashion and the accompanying computational tools are within reach. The most proximal read-outs of transcriptional activity, either DNA binding by a transcription factor or the level of the resultant transcripts, can be readily assessed on a genome-wide scale and easily subjected to digitalization. Moreover, unlike the proteome, the transcriptome currently can be examined dynamically in time course experiments over many conditions to accelerate data accession. Therefore, the first set of criteria for systems biology, dense datasets cast in a digital format, can be satisfied.

The greatest challenges in establishing this systems approach are not biological but computational and organizational. The computational issues are centered on the search and analysis of massive amounts of data (e.g., the behavior of all transcripts in a number of species over time in conditions interrogating several biological processes), on integration of heterogeneous databases (e.g., protein-protein interaction with RNA expression information with knowledge of biochemical pathways), and on large-scale data-presentation systems interpretable to bench biologists. Ultimately, the importance of any computational approach will be judged not on its mathematical beauty but by how it can be used to predict new biological phenomenon. To this end, the close physical interaction between computationally derived biological theory and experimentation will be essential for accelerating progress in the field. Traditionally, bioinformatics resided in a computer science or biostatistics department, biology in a bio-

*Correspondence: liue@gis.a-star.edu.sg

chemistry department, and a genomics center functionally and administratively disjointed from the two. This is an unfortunate disconnect and will need organizational realignment. Regarding data presentation, it is not reasonable to expect computational sophistication in an entire generation of biologists not raised in such an environment. The success of products that provide more natural interfaces between humans and computers shows the appeal (and the need) of bringing technology to the nonexpert users. Thus, we believe there will be a demand for a simplified interface specifically tailored to enable biologists to make use of such massive digitalized biological data.

Equally important for systems biology to flourish is an organizational structure and social culture that encourages collective effort and iterative experimentation between the wet laboratory and computational modeling. The current funding systems and departmental structures have a challenging task ahead. Grants for individual principal investigators (PI) are small compared to the experimental and infrastructural needs to mount significant efforts in systems biology. Thus far, grants have been administered such that incremental advances by the individual scientist, rather than an individual's contribution to larger efforts that may have much greater scientific impact, are rewarded. Funding for critical infrastructure and technology development is often bypassed as lacking scientific content. Departmental promotion procedures have tended to encourage PI insularity by discounting participation in collaborative projects focusing solely on individual effort. Graduate student training, centered on serving an individual PI's projects and demarcated by classical departmental boundaries, has not responded to the educational needs in integrative biology. Systems biologists require crossdisciplinary training that is difficult under current departmental structures.

For these reasons, we have spent a great deal of energy in crafting the cultural state of our institute, which prizes flexibility, individual intensity, and collective impact (<http://www.gis.a-star.edu.sg/homepage/>). The Genome Institute of Singapore specifically recognizes technology-focused scientists to be as important as biology-centric investigators, but we demand that technologists team with biologists to address fundamental biological questions. We devote much time to achieving consensus as to the major scientific questions we will attack together as a community. We focus on data integration so that individual efforts can be harvested for collective reasoning, though admittedly, we are still far away from the ideal. Then, our individual performance is judged not only on excellence in our personal science but also on how we contribute to the advance of our collective goals. Our project-based internal funding has allowed us as an institution to focus on the biological question rather than managing territorial claims amongst investigators. And in a project-driven culture, a postdoctoral fellow can lead a program as much as a senior group leader. We actively encourage joint mentorship for our postdoctoral fellows and graduate students, often between a technology-oriented and a biology-focused investigator or between an informatics scientist and a biologist. Without such organizational changes, the important cultural conditions for a

robust systems biology environment will not flourish. An important factor facilitating this cultural change has been block funding from a single source so that we can coordinate projects on a large scale with a long time-horizon.

The excitement over the systems approach to biology and medicine is justified. The ability to predict biological outcomes in complex systems is the grand enticement. The only limitations to our success are what we, ourselves, place before us.

References

- Bolouri, H., and Davidson, E.H. (2003). Transcriptional regulatory cascades in development: initial rates, not steady state, determine network kinetics. *Proc. Natl. Acad. Sci. USA* 100, 9371–9376.
- Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Caletani, C., Yuh, C.H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., et al. (2002a). A provisional regulatory gene network for specification of endomesoderm in the sea urchin embryo. *Dev. Biol.* 246, 162–190.
- Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Caletani, C., Yuh, C.H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., et al. (2002b). A genomic regulatory network for development. *Science* 295, 1669–1678.
- Hood, L., Heath, J.R., Phelps, M.E., and Lin, B. (2004). Systems biology and new technologies enable predictive and preventative medicine. *Science* 306, 640–643.
- Lin, C.Y., Strom, A., Vega, V.B., Kong, S.L., Yeo, A.L., Thomsen, J.S., Chan, W.C., Doray, B., Bangarusamy, D.K., Ramasamy, A., et al. (2004). Discovery of estrogen receptor alpha target genes and response elements in breast tumor cells. *Genome Biol.* 5, R66.
- Liu, E.T. (2004). Expression genomics and cancer biology. *Pharmacogenomics* 5, 1117–1128.
- Peng, X., Murthy Karuturi, R.K., Miller, L.D., Lin, K., Jia, Y., Kondur, P., Wang, L., Wong, L.S., Liu, E.T., Balasubramanian, M.K., and Liu, J. (2005). Identification of cell cycle-regulated genes in fission yeast. *Mol. Biol. Cell* 16, 1026–1042.
- Ng, P., Wei, C.L., Sung, W.K., Chiu, K.P., Lipovich, L., Ang, C.C., Gupta, S., Shahab, A., Ridwan, A., Wong, C.H., et al. (2005). Gene identification signature (GIS) analysis for transcriptome characterization and genome annotation. *Nat. Methods* 2, 105–111.
- Ruan, Y., Le Ber, P., Ng, H.H., and Liu, E.T. (2004). Interrogating the transcriptome. *Trends Biotechnol.* 22, 23–30.
- Wei, C.L., Miura, T., Robson, P., Lim, S.K., Xu, X.Q., Lee, M.Y., Gupta, S., Stanton, L., Luo, Y., Schmitt, J., et al. (2005). Transcriptome profiling of human and murine ESCs identifies divergent paths required to maintain the stem cell state. *Stem Cells* 23, 166–185.